

# Multilevel distillation of magic states for quantum computing

Cody Jones\*

Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305-4088, USA

We develop a procedure for distilling magic states used in universal quantum computing which requires substantially fewer resources than prior schemes. Our distillation circuit is based on a family of concatenated quantum codes with a transversal Hadamard operation which can distill the eigenstate of the Hadamard operator. A crucial result of this design is that low-fidelity magic states can be consumed to purify high-fidelity magic states to even higher fidelity, which we call “multilevel distillation.” When distilling in the asymptotic regime of infidelity  $\epsilon \rightarrow 0$  for each input magic state, the number of input magic states consumed on average to yield an output state with infidelity  $O(\epsilon^{2^r})$  approaches  $2^r + 1$ , which comes close to saturating the conjectured bound in [Bravyi and Haah, arXiv:1209.2426]. We show numerically that there exist multilevel protocols such that the average number of magic states consumed to distill from error rate  $\epsilon_{\text{in}} = 0.01$  to  $\epsilon_{\text{out}}$  in the range  $10^{-5}$  to  $10^{-40}$  is about  $14 \log_{10}(1/\epsilon_{\text{out}}) - 40$ ; the efficiency of multilevel distillation dominates all other reported protocols when distilling Hadamard magic states from initial infidelity 0.01 to any final infidelity below  $10^{-7}$ . These methods are an important advance for magic-state distillation circuits in high-performance quantum computing.

## I. INTRODUCTION

Quantum computing can potentially solve a handful of otherwise intractable problems, such as factoring large integers [1] or simulating quantum physics [2]. Though the number of applications with a known “quantum speed-up” is small, some are quite valuable, like the preceding examples. Quantum computations depend on coherent entangled states which are very sensitive to noise, so fault-tolerant quantum computing addresses imperfections in physical hardware with error-correcting codes [3, 4], the most studied of which are stabilizer codes [5]. However, while quantum codes protect against noise, no stabilizer code natively supports a universal set of transversal gates for simulating any quantum circuit [6]. To achieve universal quantum computing with error correction, Bravyi and Kitaev proposed a solution [7] that has received considerable attention: inject faulty “magic states” into the code, purify them using the error-corrected gates, then consume them to implement otherwise unavailable quantum circuits. These states are “magic” because it is possible to distill a subset of high-fidelity states from an ensemble of faulty states and because they enable universal fault-tolerant quantum computation.

Magic state distillation has been the subject of intense investigation in recent years. Knill independently introduced a distillation procedure for  $|H\rangle$ , the  $(+1)$  eigenstate of the Hadamard operation [8], prior to the work by Bravyi and Kitaev [7]. Reichardt showed that these protocols were equivalent and introduced an improvement which increased the threshold error rate [9]. More recently, Meier *et al.* introduced a 10-to-2 distillation procedure based on a code with two encoded qubits [10], and Bravyi and Haah introduced a  $(3k + 8)$ -to- $k$  procedure

using so-called triorthogonal codes with even  $k$  encoded qubits [11]. The distillation procedures we develop herein continue this trend of using larger, multi-qubit codes. Additionally, Fowler and Devitt have proposed methods to reduce the overhead for distillation circuits when using topological quantum error correction [12].

For a quantum state, we quantify the probability of it having an error using the infidelity  $1 - F$ , where  $F = \langle \psi | \rho | \psi \rangle$  is the fidelity between some mixed state  $\rho$  and the ideal state  $|\psi\rangle$ . The initial  $|H\rangle$  states are prepared in a faulty manner before being injected into a fault-tolerant quantum code, and Reichardt proved that the theoretical-maximum infidelity for  $|H\rangle$  states to be distilled is about 0.146 [9]. The efficiency of magic state distillation is of great importance to fault-tolerant quantum computing. Although magic states are the widely preferred method for achieving universality, distillation circuits are currently estimated to require the majority of resources in a quantum computer [13, 14]. Therefore, advances in distillation protocols are important steps toward making quantum computing possible.

This paper presents two important, related results. First, we specify a family of  $[[n, (n - 4), 2]]$  Calderbank-Shor-Steane (CSS) quantum stabilizer codes [15, 16] known as “ $H$  codes” with transversal Hadamard operation, for even  $n \geq 6$ . These codes are dense, in the sense that the ratio of logical qubits to physical qubits  $(n - 4)/n \rightarrow 1$  as  $n \rightarrow \infty$ . The  $H$  codes allow distillation of magic states with the same  $(3k + 8)$ -to- $k$  efficiency as Ref. [11], and later we present further evidence that there may be connections between these methods. Second, we demonstrate that concatenated versions of  $H$  codes allow for distillation of high-fidelity encoded magic states by consuming low-fidelity magic-state ancillas. We call this “multilevel distillation,” and it leads to the most efficient procedure for distilling magic states reported so far. For suitably small infidelity  $\epsilon$  in each input magic state with independent errors, there exists a multilevel distillation protocol that yields output magic states with

\* ncodyjones@gmail.com

infidelity  $O(\epsilon^{2^r})$  and requires asymptotically  $2^r + 1$  input states per distilled output state. This efficiency comes close to the “optimality” bound conjectured in Ref. [11]. While this result is interesting theoretically, we also numerically study the distillation efficiency for  $\epsilon_{\text{in}} = 0.01$ , which is of practical importance to fault-tolerant quantum computing. We find that multilevel distillation is superior to previously reported protocols when the final fidelity is below  $10^{-7}$ .

Throughout this paper, we adopt the following notation for single-qubit Pauli operators, for readability:  $X \equiv \sigma^x$ ,  $Z \equiv \sigma^z$ , and  $I$  is the identity operator on a qubit. Additionally, we use “physical qubit” to denote those qubits used to produce a quantum code (including the redundancy of code stabilizers [4, 5]), whereas “logical qubits” are the logical degrees of freedom inside the code, again for readability. It may be the case that physical qubits are themselves the logical qubits of another code, providing fault tolerance, which is a standard technique of quantum code concatenation [3, 4, 17].

## II. A FAMILY OF CODES WITH TRANSVERSAL HADAMARD

We define a family of CSS quantum codes which encode an even number  $k$  logical qubits using  $(k + 4)$  physical qubits and possess a transversal Hadamard operation, so we call them collectively “ $H$  codes” and denote  $H_n$  as the code using  $n = k + 4$  physical qubits. The transversal Hadamard operation at the physical level enacts a transversal Hadamard operation at the logical level, which will be a useful property when we later concatenate these codes. Table I specifies the stabilizers [4, 5] and logical operators for any  $H$  code. The smallest code,  $H_6$ , is shown explicitly, and other codes can be generated from a simple pattern, as follows. Using our ordering, let us label the first four physical qubits as the “preamble,” while the remaining physical qubits are “index” qubits. The operators in the preamble are the same for all  $H$  codes. After the preamble, the stabilizers have an  $I$ ,  $X$ , or  $Z$  on every index qubit. The  $i^{\text{th}}$  logical  $\bar{X}_i$  has an  $X$  acting on the  $i^{\text{th}}$  index qubit, and  $I$  on all other index qubits; the logical  $\bar{Z}_i$  follows the same pattern.

It is straightforward to verify that the  $H$  codes have a transversal Hadamard operation. Application of Hadamard gates to all physical qubits simply interchanges stabilizer generators, while it maps  $\bar{X}_i \mapsto \bar{Z}_i$  and  $\bar{Z}_i \mapsto \bar{X}_i$ . Additionally, these are distance-2 codes because the Kronecker product of two logical Pauli operators of the same type for two distinct logical qubits has weight 2 (physical Pauli operators); the Kronecker product of same-type Pauli operators on all logical qubits is also weight-2. The stabilizers come in matched  $X/Z$  pairs, so there are no weight-1 logical operators.

The  $(+1)$  eigenstate  $|H\rangle$  of the Hadamard operator  $H = (1/\sqrt{2})(X + Z)$  is a magic state for universal quan-

Stabilizers	XXXX II(I...)
	ZZZZ II(I...)
	XIXI XX(X...)
	ZIZI ZZ(Z...)
$\bar{X}_1 =$	XXII XI(I...)
$\bar{X}_2 =$	XXII IX(I...)
	(etc.)
$\bar{Z}_1 =$	ZZII ZI(I...)
$\bar{Z}_2 =$	ZZII IZ(I...)
	(etc.)

TABLE I. Stabilizer generators and logical operators for the family of  $H$  codes. The notation XXXX is shorthand for a Kronecker product of the Pauli operator  $\sigma^x$  on four qubits, where order indicates which qubits these operators act on. The smallest code  $H_6$  is shown, and the operators in parentheses show how to extend the pattern to larger  $H$  codes.

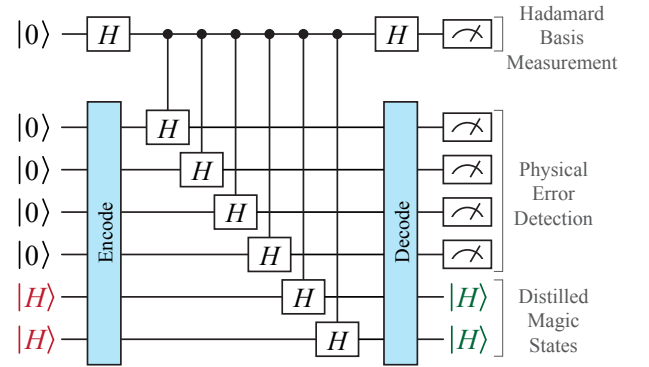


FIG. 1. Distillation of  $|H\rangle$  magic states using the  $H_6$  code. Two initial  $|H\rangle$  states (red) are encoded with four additional qubits, initialized to  $|0\rangle$  here. The boxes “Encode” and “Decode” represent quantum circuits for encoding and decoding the  $H_6$  code, which are not shown here. The procedure succeeds if neither the Hadamard-basis measurement nor the error-detection measurement indicate an error, and the result is two  $|H\rangle$  states of higher fidelity; otherwise, the results are discarded. Each controlled-Hadamard gate consumes two  $|H\rangle$  states, so this is a 14-to-2 distillation procedure.

tum computing [7–11]. In particular, two of these magic states can be consumed to implement a controlled- $H$  operation [8, 10], enabling one to measure in the basis of  $H$ . Our distillation procedure is based on the following heuristic: (a) encode faulty  $|H\rangle$  magic states in an  $H$  code; (b) measure in the basis of the transversal Hadamard gate by consuming  $|H\rangle$  ancillas; (c) reject the output states if either the measure-Hadamard or code-stabilizer circuits detect an error. For example, when an  $H_{(k+4)}$  code is used for distillation,  $k$   $|H\rangle$  states are encoded as logical qubits using  $(k+4)$  physical qubits. Each transversal controlled-Hadamard gate consumes two  $|H\rangle$  states [10], and this gate is applied to all physical qubits, which leads to the  $(3k + 8)$ -to- $k$  input/output distillation efficiency of these codes. A diagram of the quantum circuit for distillation using  $H_6$  is shown in Fig. 1.

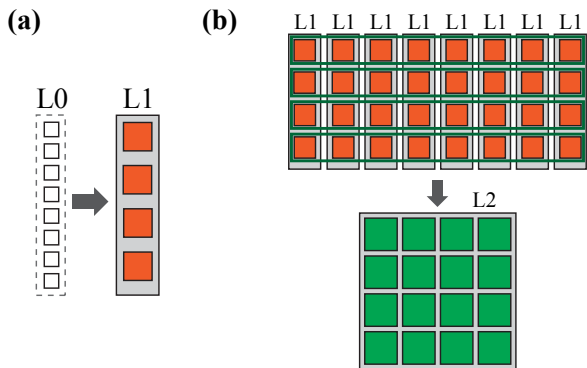


FIG. 2. Concatenation of  $H$  codes. (a) Eight bare “L0” qubits are encoded in an  $H_8$  code to form four logical “L1” qubits. (b) The L1 logical qubits from eight copies of the  $H_8$  code are then encoded at another level with  $H_8$  codes, yielding a  $[[64, 16, 4]]$  concatenated code with 16 logical “L2” qubits.

### III. MULTILEVEL DISTILLATION

Multilevel distillation uses concatenated codes with transversal Hadamard for distillation, in such a manner that the  $|H\rangle$  ancillas consumed for transversal controlled-Hadamard measurement are of lower fidelity than the encoded logical  $|H\rangle$  states being distilled. When two quantum codes with transversal Hadamard are concatenated, the resulting code also has transversal Hadamard. Under appropriate conditions, the distance of the concatenated code is the product of the distances for the individual codes:  $d' = d_1 d_2$  [10]. Specifically, the logical qubits of the first level of encoding are used as physical qubits for completely distinct codes at the second level. Consider a two-level scheme: if the codes at first and second levels are  $[[n_1, (n_1 - 4), 2]]$  and  $[[n_2, (n_2 - 4), 2]]$ , respectively, then the concatenated code is  $[[n_1 n_2, (n_1 - 4)(n_2 - 4), 4]]$ , as shown in Fig. 2. Thus the concatenation of two  $H$  codes yields a distance-4 code with transversal Hadamard, and  $r$ -level concatenation has distance  $2^r$ .

At the logical level, the transversal controlled-Hadamard operation detects a single error in the encoded qubits, for any degree of concatenation. If the logical  $|H\rangle$  states have independent error probabilities  $\epsilon_1$ , then the distilled states will have infidelity  $O(\epsilon_1^2)$  with perfect distillation. In the preceding example, the transversal controlled-Hadamard gates at the lowest physical level require  $(2n_1 n_2)$   $|H\rangle$  magic states, each of which has infidelity  $\epsilon_2$ . However, this is a distance-4 code, so for independent input error rates, the probability of failing to detect errors at the physical level is  $O(\epsilon_2^4) + O(\epsilon_1 \epsilon_2^2)$  (rigorous analysis is provided below). Therefore, the  $|H\rangle$  states consumed at the physical level can be of lower fidelity than the magic states encoded as logical qubits and successfully perform distillation.

One type of error event is not detected by either the controlled-Hadamard measurement or the error-

detection circuits of the code. If there is an error in an encoded magic state and errors on two physical states used for the same controlled-Hadamard gate at the physical level, then the combination of these errors will cause a faulty Hadamard-basis measurement while not being detected by the  $H$  code. This event leads to the  $O(\epsilon_1 \epsilon_2^2)$  error probability from before, which is not an issue at levels one and two, but it must be addressed in levels three and up. The solution for level- $r$  distillation, where  $r \geq 3$ , is to repeat the controlled-Hadamard measurement  $2^{(r-2)}$  times, consuming  $2^{(r-1)}$  magic states at the physical level. After each transversal controlled-Hadamard, the code syndrome checks for detectable error patterns. With this procedure, one encoded-state error would also require at least  $2^{(r-1)}$  errors in physical-level magic states to go undetected.

Multilevel distillation circuits tend to be much larger in both qubits and gates than other protocols. Because there can be many encoded qubits, the protocol is still very efficient, but the size of the overall circuit is large. At any number of levels, the distilled output states have correlated errors, so distilled magic-state qubits in our protocol must never meet again in a subsequent distillation circuit (we require that errors are independent within the same encoding block, as in Ref. [10]). Let us suppose that one performs two rounds of distillation, where the first round uses one-level distillers with  $k$  encoded magic states and the second round uses two-level distillers with  $k^2$  encoded states. Because the inputs to the second round must have independent errors, the total number of magic states required is  $5k^3 + O(k^2)$ . For  $r$  rounds in this fashion, the total number of input magic states is  $(2^r + 1)(k)^{r(r+1)/2} + O((k)^{[r(r+1)/2]-1})$ , so the overall size of the procedure becomes unwieldy after three or four rounds. However, since efficient multilevel distillation protocols, measured in the ratio of low-fidelity  $|H\rangle$  input states consumed to yield a single high-fidelity  $|H\rangle$  output, use  $k \gg 1$  and multiple rounds, the greatest benefit from their application is seen in large-scale quantum computing, where a typical algorithm run may require  $10^{12}$  magic states, each with error probability  $10^{-12}$  [14].

The “scaling exponent”  $\gamma$  of a distillation protocol characterizes its efficiency. Specifically,  $O(\log^\gamma(\epsilon_{\text{in}}/\epsilon_{\text{out}}))$  input states are required to distill one magic state of infidelity  $\epsilon_{\text{out}}$ . Scaling exponents for previous protocols are  $\gamma \approx 2.46$  (“15-to-1” [7, 8]),  $\gamma \approx 2.32$  (“10-to-2” [10]), and  $\gamma \approx 1.6$  (triorthogonal codes [11]). Moreover, Bravyi and Haah conjecture that no magic-state distillation protocol has  $\gamma < 1$  [11]. In this work, each increasing level of multilevel distillation doubles the number of consumed inputs because of the repeated controlled-Hadamard measurements. In the limits of  $k \rightarrow \infty$ ,  $\epsilon \rightarrow 0$ , multilevel protocols consume  $2^r + 1$  input states to each output state for  $r$  rounds of distillation. The final infidelity is  $O((\epsilon_{\text{in}})^{2^r})$ , so the scaling exponent is  $\gamma = \log(2^r + 1)/\log(2^r) \rightarrow 1$  as  $r \rightarrow \infty$ , which is the closest any protocol has come to reaching the conjectured bound. We show below that  $\gamma \approx 1$  for error rates relevant to quantum computing.

#### IV. ANALYSIS

We make the conventional assumption that all quantum circuit operations are perfect, except for the initial  $|H\rangle$  magic states we intend to distill; for a more explicit treatment, see Ref. [12]. This is a valid approximation if all operations are performed using fault-tolerant quantum error correction where the logical gate error is far below the final infidelity for distilled magic states [3, 14]. Additionally, following the methodology in Refs. [7, 10], we can consider each magic state with infidelity  $\epsilon$  as the mixed state  $(1 - \epsilon)|H\rangle\langle H| + \epsilon|-H\rangle\langle -H|$ , where  $|-H\rangle$  is the  $(-1)$  eigenstate of the Hadamard operation.

Determining the infidelity at the output of distillation becomes simply a matter of counting the distinct ways that errors lead to the circuit incorrectly accepting faulty states. Denote error probabilities for encoded states as  $\epsilon_1$  and for states consumed for physical-level controlled Hadamard as  $\epsilon_2$ , which are all assumed to be independent. Then a one-level,  $(3k + 8)$ -to- $k$  distiller using the  $H_{(k+4)}$  code has output error rate on each  $|H\rangle$  state as

$$\epsilon_{\text{out}} = (k - 1)\epsilon_1^2 + (2k + 2)\epsilon_2^2 + \dots, \quad (1)$$

where higher order terms denoted (...) are omitted since we are interested in the  $\{\epsilon_1, \epsilon_2\} \rightarrow 0$  regime. The lowest-order error rates are both second order, because the Hadamard basis measurement and  $H_{(k+4)}$  code can together detect a single error in any magic state. The probability of the distiller detecting an error, in which case the output is discarded, is  $k\epsilon_1 + 2(k + 4)\epsilon_2 + O(\epsilon_1^2) + O(\epsilon_2^2)$ . If  $\epsilon_1 = \epsilon_2 = \epsilon$ , then the output error rate of  $(3k + 1)\epsilon^2$  conditioned on success is the same as in Ref. [11].

The two-level distiller constructed from concatenated  $H_{(k+4)}$  codes can detect any of: a single error in the  $k^2$  encoded magic states; at least three errors in the  $2(k + 4)^2$   $|H\rangle$  states consumed at the lowest physical level; or one encoded-state error and one physical error. The output error rate for each distilled  $|H\rangle$  state is

$$\epsilon_{\text{out}} = (k^2 - 1)\epsilon_1^2 + 8(k^2 + 4k + 3)\epsilon_2^4 + (k + 4)\epsilon_1\epsilon_2^2 + \dots \quad (2)$$

The probability of the two-level distiller detecting an error is  $k^2\epsilon_1 + 2(k + 4)^2\epsilon_2 + 2k^2(k + 4)^2\epsilon_1\epsilon_2 + O(\epsilon_1^2) + O(\epsilon_2^2)$ . Multilevel distillation for level  $r \geq 3$  requires repeating the controlled-Hadamard measurement  $2^{r-2}$  times. Doing so ensures that the lowest-order error events are:  $O(\epsilon_1^2)$  for two encoded errors;  $O((\epsilon_2)^{2^r})$  for  $2^r$  physical-level errors; and  $O(\epsilon_1(\epsilon_2)^{2^{(r-1)}})$  for one encoded error and  $2^{(r-1)}$  physical errors that come in pairs not detected by the code, but propagate to all of the Hadamard-basis measurements. If  $\epsilon_1 \sim (\epsilon_2)^{2^{(r-1)}}$ , then all three of these lowest-order error events are approximately equally likely, and the protocol is balanced.

Figure 3 shows the performance of distillation protocols identified by numerical search. The markers indicate the type of the last round of distillation. As expected, there is a trend of using higher-distance multilevel protocols as the output error rate  $\epsilon_{\text{out}}$  decreases. The plot also

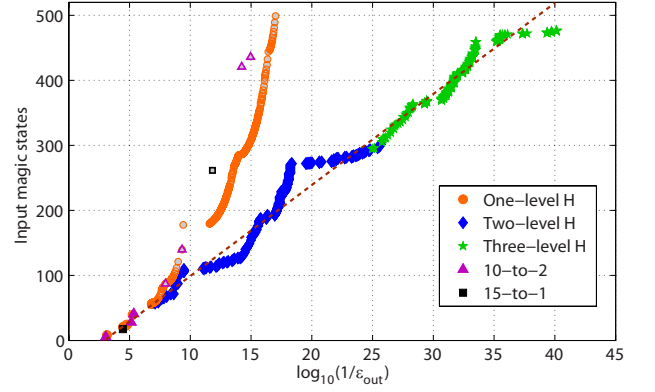


FIG. 3. Average number of input  $|H\rangle$  states with  $\epsilon_{\text{in}} = 0.01$  consumed to produce a single output  $|H\rangle$  state with fidelity  $\epsilon_{\text{out}}$ . Multiple-round distillation can use different protocols in each round. The markers indicate the protocol type for the last round of distillation. Only efficient protocols are plotted, meaning that for a given  $\epsilon_{\text{out}}$ , no other protocol requires fewer input states; the search space was constrained such that encoded qubits  $k \leq 40$  for all  $H$  codes and number of rounds  $r \leq 5$ . The grey-shaded squares, triangles, and circles show, respectively, the best distillation with 15-to-1 [7], 10-to-2 [10], and triorthogonal-code [11] protocols. The dashed line is a linear fit  $14 \log_{10}(1/\epsilon_{\text{out}}) - 40$ .

shows the best results achievable with prior protocols, and multilevel distillation is dominant for  $\epsilon_{\text{out}} \leq 10^{-7}$ , which is the regime pertinent to quantum computing. The linear fit provides empirical evidence that the scaling exponent is  $\gamma \approx 1$  in this regime.

#### V. CONCLUSIONS

Interesting similarities between the work in Ref. [11] and our results merit further study. The distillation ratio  $(3k + 8)$ -to- $k$  and the lowest-order error of  $(3k + 1)\epsilon^2$  for the  $H$  codes are exactly the same as for distillation using triorthogonal codes [11]. Moreover, Reichardt has previously shown a connection [9] between distillation using the Reed-Muller code with transversal phase gate  $T = \exp(i\pi\sigma^z/8)$ , like the triorthogonal codes, and distillation using the Steane code with transversal Hadamard, like our  $H$  codes; this suggests a similar connection may exist between the work herein and that of Ref. [11]. A high-distance, high-density CSS code with transversal  $T$  would be desirable since it would allow one to directly distill a three-qubit magic state for controlled-controlled- $Z$ , which is locally equivalent to the Toffoli gate [4].

Multilevel distillation is an important development for large-scale, fault-tolerant quantum computing, where the distillation of magic states is often considered the most costly subroutine [13, 14]. Other codes with high density, high distance, and transversal Hadamard may yet be discovered, though for the present,  $H$  codes are useful for their high efficiency and simple construction.

## ACKNOWLEDGMENTS

The author would like to thank Sergey Bravyi and Bryan Eastin for assistance in analyzing protocols at levels three and higher. This work was supported by the

Univ. of Tokyo Special Coordination Funds for Promoting Science and Technology, NICT, and the Japan Society for the Promotion of Science (JSPS) through its “Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program).”

- 
- [1] P. W. Shor, SIAM J. Comput. **26**, 1484 (1997).
  - [2] S. Lloyd, Science **273**, 1073 (1996).
  - [3] J. Preskill, P. Roy. Soc. A-Math. Phys. **454**, 385 (1998).
  - [4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 1st ed. (Cambridge University Press, 2000).
  - [5] D. Gottesman, *Stabilizer codes and quantum error correction*, Ph.D. thesis, California Institute of Technology (1997).
  - [6] B. Zeng, A. Cross, and I. Chuang, Information Theory, IEEE Transactions on **57**, 6272 (2011).
  - [7] S. Bravyi and A. Kitaev, Phys. Rev. A **71**, 022316 (2005).
  - [8] E. Knill, “Fault-tolerant postselected quantum computation: Schemes,” (2004), *Preprint* arXiv:quant-ph/0402171.
  - [9] B. W. Reichardt, Quantum Information Processing **4**, 251 (2005).
  - [10] A. M. Meier, B. Eastin, and E. Knill, “Magic-state distillation with the four-qubit code,” (2012), *Preprint* arXiv:1204.4221v1.
  - [11] S. Bravyi and J. Haah, “Magic state distillation with low overhead,” (2012), *Preprint* arXiv:1209.2426v1.
  - [12] A. G. Fowler and S. J. Devitt, “A bridge to lower overhead quantum computation,” (2012), *Preprint* arXiv:1209.0510v3.
  - [13] N. Isailovic, M. Whitney, Y. Patel, and J. Kubiawicz, in *35th International Symposium on Computer Architecture, 2008 (ISCA’08)* (2008).
  - [14] N. C. Jones, R. Van Meter, A. G. Fowler, P. L. McMahon, J. Kim, T. D. Ladd, and Y. Yamamoto, Phys. Rev. X **2**, 031007 (2012).
  - [15] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1098 (1996).
  - [16] A. Steane, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **452**, 2551 (1996).
  - [17] E. Knill and R. Laflamme, “Concatenated quantum codes,” (1996), *Preprint* arXiv:quant-ph/9608012.